**SI 1.** Summary of the percentage of missing values present in four datasets and the correlation of missing values observed with *m/z*, retention time and response. Filtering was performed as defined in section 2.1.1.3.

| Processing method | Methods applied (number of papers reported) |
|---|---|
| Normalisation | PQN (2), normalization to IS (9), SUM (3), mean/median normalisation (3), other (3), not reported (31) |
| Missing value imputation | KNN (1), mean replacement (1), small value replacement (3), not reported (46) |
| Transformation | Natural logarithm (12), $Log_2$(5), $Log_{10}$(1), other (1), not reported (32) |
| Scaling | Autoscaling (11), Pareto scaling (6), level scaling (1), not reported (33) |